

Gaze-contingent audio-visual substitution for the blind and visually impaired

Lukas Twardon, Hendrik Koesling, Andrea Finke and Helge Ritter
Center of Excellence Cognitive Interaction Technology - CITEC
Bielefeld University
Bielefeld, Germany
{ltwardon, ihkoesli, afinke, helge}@techfak.uni-bielefeld.de

Abstract—A lot of effort has been made to advance technologies that improve blind and partially sighted people’s spatial perception. One common approach is to enhance or substitute vision by audition. Most sensory substitution systems, however, have not attached any importance to eye movements. But eye movements play an essential role in mental imagery even in the absence of visual input. Therefore, we propose a system for gaze-contingent auditory substitution of spatial vision. It is intended to be a mobile helper in everyday life of the visually impaired. The prototype we have developed combines eyetracking with depth measuring and sonification techniques. We carried out both a proof-of-concept study in complete darkness and an exploratory EEG study. Our findings indicate that gaze-contingent sensory substitution permits depth perception and leads to intermodal (audio-visual) processing in untrained subjects. Hence, as a result of neuroplasticity, the blind and visually impaired might learn to perceive gaze-dependent sound visually.

I. INTRODUCTION

Sighted people are used to perceiving depth mostly visually. Humans, in contrast to echolocating species like bats, are typically much better at processing the binocular retinal image than at measuring time differences between acoustic waves. Technical systems, however, are capable of sensing physical quantities that are otherwise inaccessible to the human brain. Infrared depth cameras, for instance, record spatial distances even in complete darkness. A converter that translates the measured values into an appropriate acoustic signal could make it possible for humans to hear spatiality. Thus, pervasive computing technologies can help in perceiving the world in a completely new way.

The transformation of one sensory modality into another one is often referred to as sensory substitution. Blind and visually impaired people particularly benefit from sensory substitution systems. Several researchers have aimed to make their subjects “see” with sound [1] [2] [3] or by touch [4] [5]. In most cases, a video camera image has been the visual source. We suggest acoustically presenting depth data rather than color or brightness because spatial perception is an important requirement for orienting and navigating.

Eye movements support mental imagery regardless of the presence of visual stimuli [6] [7] [8]. Therefore, we have developed the *Auditory Night Sight (ANS)* as a gaze-contingent system, meaning that the user’s gaze immediately determines the direction of perception. The *ANS* basically consists of a depth camera, an eyetracker and a sound generator. A low

pitch sound is produced when gaze is directed to a far away object, a short distance in gaze direction leads to a high pitch sound.

II. BACKGROUND

A. Multisensory integration and brain plasticity

The mental representation of the world arises from the interaction of various modalities, including vision and audition. The brain reliably matches visual and auditory signals originating from the same source [9] although the modalities differ a lot. Multisensory integration is context-dependent. That is, information coming from one sense influences the mental image more or less depending on how reliable the data are. Alais and Burr [10] showed that mainly visual cues are used to localize an audio-visual stimulus if vision is unrestricted. For blurred visual stimuli, in contrast, sound dominates vision. But what if one sensory modality is missing or limited for a long period of time, as is the case with blindness and visual impairment? Then, neuroplasticity comes into play.

Neuroplasticity is the capacity of the brain to change its own structure and function [11]. This principle is assumed to be the neural basis for learning [12]. Plasticity can occur both on a synaptic level and on a level concerning whole cortical brain areas. The latter is referred to as cortical plasticity. Constraint-induced movement therapy tries to take advantage of cortical plasticity for rehabilitation purposes [13]. For example, the right arm can intentionally be immobilized in order to improve the cortical control over the left arm. The fact that the increased use of a body part leads to a neural reorganization process could also be beneficial in our case. There is hope that gaze-contingent acoustic stimulus presentation improves even blind people’s eye movement control. Furthermore, the close link between eye movements and visual perception (see section II-C) could induce brain plasticity. Gaze-dependent auditory stimuli might be processed in traditionally visual brain areas after a learning phase.

B. Sensory substitution

How can technical systems support neuroplasticity, and is technology even necessary? It is well known that, in most cases, the blind localize sound sources better than sighted people [14]. Most obstacles, however, do not emit any sound

and are therefore hard to detect without vision. That is why some blind people have learned to echolocate. They produce certain clicking sounds with their tongue and interpret the echoes. In this way, they can determine the position and shape of objects [15]. Thaler et al. [16] found that human echolocation experts process the echoes mostly in visual brain areas.

One might conclude that sensory substitution works perfectly without any technical assistance. But human echolocation is hard to learn, and there are only very few people who can do it. Therefore, many devices with electronic sensors and stimulators have been developed. The substituted sense is not always vision, but also the vestibular system [17] and even spatial orientation obtained by a magnetic compass [18]. The measured data has been presented to the user either tactilely or acoustically. There is also the idea to have invasive stimulators [19]. However, implants have the disadvantage of being irreversible.

Self-induced motion is crucial for spatial perception in sensory substitution. The sensory substitution pioneer Paul Bach-y-Rita [20] noticed: "*Self-induced camera motion appears to be analogous to eye movements. [...] a translation of the input that is precisely correlated with self-generated movement is the necessary and sufficient condition for the experienced phenomena to be attributed to a stable outside world.*" Since the input is supposed to be correlated with motion analogous to eye movements, our approach is to actually use eye movements.

C. Eye movements and mental imagery

Eye movements play an important role in visual perception. One essential function is the fixation of objects whereby the fovea is directed towards the stimulus. According to the eye-mind hypothesis, eye movements and attention are highly correlated which was also shown on a neural basis [21]. But does the role of eye movements in mental imagery go beyond focusing an object?

It makes sense to have a look at the process of recalling visual information. If focusing is the only function of eye movements, they should not occur during the recall process. But Brandt and Stark [6] showed that eye movement patterns were nearly the same whether a scene was actually viewed or just visualized. Laeng and Teodorescu [7] confirmed the results and even found that subjects who were not allowed to move their eyes during the visualization of a previously perceived scene, showed decreased recall abilities. The authors conclude that commands to the eyes are stored as spatial indices in order to properly arrange partial images. Furthermore, Johansson et al. [8] have revealed that eye scanpaths during visualization reflect the positions of objects, regardless of whether the scene was visually perceived or orally described. Also, the results were the same in complete darkness. Thus, eye movements are crucial for mental imagery, even if vision is not involved.

III. RELATED WORK

Over the past decades, many systems have been developed that try to compensate blindness or visual impairment. We will only discuss a few of them which are most relevant to us. Most systems can roughly be separated into two categories. On the one hand there are conventional assistive systems that do a lot of filtering and preprocessing on the recorded data in order to present to the user only the information which is needed to fulfill a certain task. The task may, for example, be navigation or obstacle avoidance. On the other hand sensory substitution systems try to induce cortical plasticity by directly transforming the visual input into a signal of another modality.

NAVI (Navigational Aid for Visually Impaired) is an assistive system that has been developed by students from the University of Konstanz [22]. What it has in common with the *ANS* is the fact that it uses the *Kinect* depth camera. A pulsed vibration signal on the left, right or center of the user's waist indicates a detected obstacle. Furthermore, marker detection is applied for global speech synthesis navigation.

Bach-y-Rita's *TVSS (Tactile-visual Substitution)* has been one of the first sensory substitution devices [4]. The system consists of a video camera which the user controls, a signal converter and a chair with a 20 x 20 matrix array of teflon tips fixed at its back. The teflon tips vibrate against the back of the user depending on the grayscale value of the camera at the correspondent position. Subjects were able to detect complex patterns after 20 to 40 hours of training. The *Teletact* device is another tactile-visual substitution system [5]. Instead of recording a grayscale image, a simple telemeter measures distances. The user just points in a certain direction and gets force feedback indicating the measured depth value. Thus, user-induced motion is realized through turning the wrist.

The *vOICE* project (OIC stands for "Oh I see") has been initiated by Meijer [1]. Both pixel position and brightness of a video camera image are coded acoustically. The image is scanned column by column which takes one second. Sound pitch is related to the vertical pixel position, amplitude is a function of pixel brightness. While the scan-based *vOICE* system has a low temporal resolution, *PSVA (Prosthesis Substituting Vision for Audition)* is a real-time auditory substitution system [2]. The transformation of the camera image into an acoustic representation is based on a model of human vision and an inverse model of human audition. A complex algorithm translates the whole image into sound all at once. The *SeeColOr* system even converts color and depth information from a stereo camera to an acoustic signal [3]. All described auditory substitution systems, however, lack in user-induced motion. Hence, the unique feature of the *ANS* system is the gaze-contingency approach.

IV. AUDITORY NIGHT SIGHT (ANS)

We have developed a prototype of a gaze-contingent audio-visual substitution system. It even works in darkness as it is infrared-based. Therefore, we named the system *Auditory Night Sight (ANS)*. The appearance is rather functional because the device has been developed only for experimental purposes.

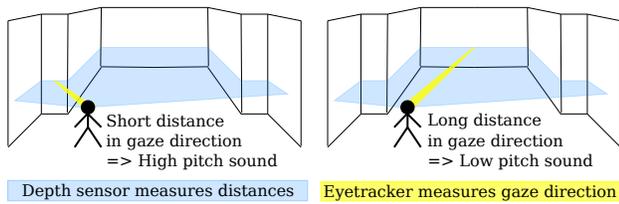


Fig. 1. Functionality of the *Auditory Night Sight (ANS)* from a user's perspective. A sine tone is played via earphones, pitch coding the measured distance (depicted in blue) in gaze direction (yellow).

The *ANS* is intended to help the user orient and navigate. Figure 1 shows the system from a user's perspective. The working principles of the *ANS* are the following.

- **Substitution of spatial vision:** Many sensory substitution systems use a conventional video camera as a sensor. However, color and brightness information is secondary for orientation and navigation tasks. Depth information, in contrast, is essential for spatial perception.
- **Auditory stimulation:** It would have been possible to project depth data as a vibration signal onto the user's skin. However, the human brain is used to perceiving spatial distances audio-visually, while touch is not naturally involved. The current version of the *ANS* transforms exactly one depth value into a sine tone of a certain frequency at a specific point in time.
- **Gaze-contingency:** The user's eye movements continuously determine the direction of acoustic depth coding. Thus, the requirement to have self-induced motion in sensory substitution is fulfilled. Eye movements are a good choice because of their important role in mental imagery. Eye position control in the dark is possible [23]. Late blind people are able to make directed eye movements [24]. Early blind individuals might improve their eye position control abilities through acoustic feedback.
- **Real-time coding:** Eye movements are very fast. Therefore, it is important to generate an auditory response immediately. Pitch changes are performed in real-time in order to create a feedback loop between the human brain and the *ANS* system.

Blind and visually impaired people are our focus group, but sighted people could also benefit from gaze-contingent sensory substitution research. Using the *ANS* can be learned with an acceptable amount of time since sonification is kept quite simple. This could be a motivation to make use of auditory substitution in various situations. One might think of, for example, rescue operations in surroundings where vision is restricted. But for all that, the *ANS* currently is a scientific tool that helps us study the role of eye movements in auditory perception, brain plasticity or sonification of depth data.

The prototype is a head-mounted system. The main hardware components are a tiltable depth camera, an eyetracker, an accelerometer and earphones. An aluminum mount joins eyetracker and depth camera together. The whole system setup is depicted in Figure 2. We have implemented a graphical user

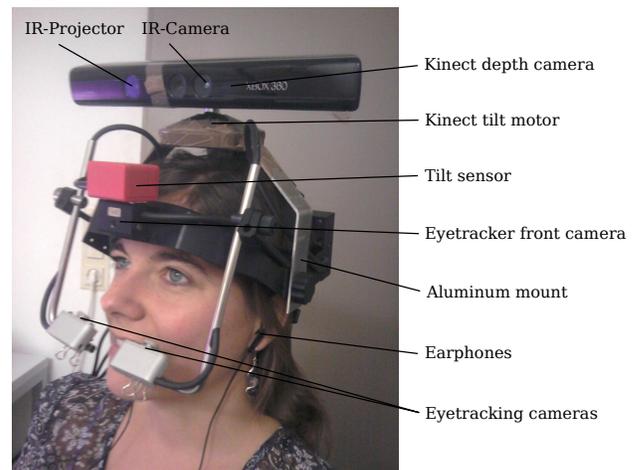


Fig. 2. Hardware components of the *ANS*. The system is mounted on the user's head.



Fig. 3. Screenshot of the *ANS* graphical user interface. The depth image is depicted as a grayscale image from black ($\leq 0.5m$) to white ($\geq 4.5m$). The current gaze point is plotted in red. Available keyboard commands are listed below.

interface including a mouse simulation mode and the option to reinitialize the system (see Figure 3).

Distance measuring is accomplished by a *Kinect*¹ depth camera. The underlying principle is the structured light technique [25]. An infrared light pattern is emitted by the projector. The infrared camera then records the reflected pattern, so that a depth image can be triangulated from differences between the original and the grabbed pattern. Stereo recording and time-of-flight measuring are alternative depth sensing approaches. But Pece et al. [26] compared the *Kinect* camera with both a stereo camera and a time-of-flight camera and have found that it can keep up with more expensive systems in terms of accuracy and resolution. Also, it works in complete darkness. We have used the open source library *libfreenect*² to access the depth image.

¹<http://www.xbox.com/kinect>

²<http://www.openkinect.org>

Eyetracking means continuously capturing the user’s gaze point and eye movements, respectively. It is very popular as an experimental method in several research areas because it is commonly assumed that there is a strong connection between eye movements and attention. Eyetracking data are usually analyzed offline, that is to say, after having finished the experiment. The *ANS*, in contrast, uses the data online in order to determine the depth value to sonify. There are many different eyetracking methods [27]. We chose the *Eyelink II*³ eyetracker which is infrared camera-based. The head-mounted system is preferably applied in screen experiments but can also be configured for mobile use. The *JoyWarrior24F14*⁴ accelerometer is used as a tilt sensor in order to be able to align the depth camera with the eyetracker’s coordinate system.

We have specified a rather straightforward sonification rule. Spatial distance is mapped to the pitch of a sine tone. Let $v_{min}(= 0.5m)$ be the minimum and $v_{max}(= 4.5m)$ be the maximum depth value. Then the frequency f is given as a function of the current depth value v by

$$f = 220 \cdot 2^{\frac{v_{max}+v_{min}-2v}{v_{max}-v_{min}}} [Hz], \text{ where } v_{min} \leq v \leq v_{max}. \quad (1)$$

The exponential mapping is contrary to the logarithmic perception of frequency in human audition. Therefore, the user perceives a linearity between distance and pitch. One could think of more sophisticated sonification approaches but we considered this as an overhead for our experimental purposes, as we have focused on the key principles of gaze-contingent sensory substitution.

V. EVALUATION IN THE DARK

A. Hypotheses

We made a proof-of-concept study in a completely darkened room in order to find out if, and how well, the *ANS* system supports untrained users’ spatial perception. Also, we tested the influence of acoustic noise on the subjects’ performance. The hypotheses were the following:

- 1) Using the *ANS*, untrained subjects can localize and determine the size of objects even in complete darkness.
- 2) Acoustic background noise does not have a negative effect on the perception of the objects.
- 3) Subjects move their eyes rather than their head to solve the task.

B. Experimental setup and method

We conducted the experiment with ten sighted untrained subjects (seven male, three female). After a written and oral explanation the subject put on the *ANS* helmet and the system was initialized and calibrated. Then, the subject could familiarize himself with the system. During the trials the subject sat in front of a $2m \times 2m$ canvas, in a distance of $1.30m$ (see Figure 4). There was about $2m$ free space behind the canvas. The canvas had four windows that could either be

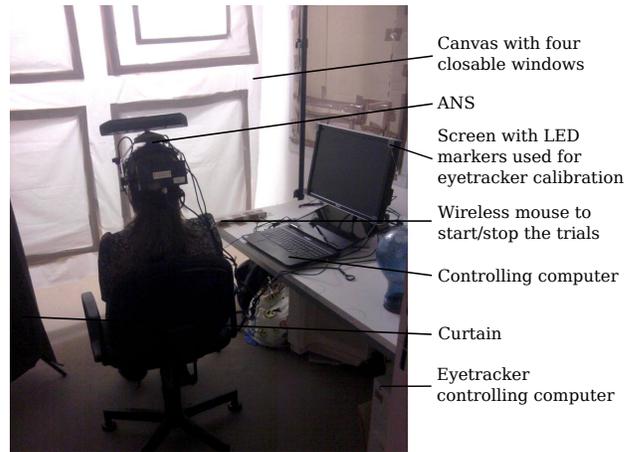


Fig. 4. Experimental setup of the proof-of-concept study which took place in complete darkness.

closed, opened in a range of $0.5m \times 0.5m$ (small window) or opened in a range of $0.75m \times 0.75m$ (large window).

Before each of the 16 trials a curtain was drawn between subject and canvas, and music was played in order to distract the subject. The experimenter opened either a small or a large window and closed the other windows. Then, light was switched off and the curtain was opened. The room was darkened in such a way that the subject had no visual perception at all. The subject performed a mouse click in order to start the trial and the gaze-contingent sonification of the depth image. We varied not only the window size but also the presence of distracting street noise. In the street noise condition, the root mean squares of depth sonification signal and noise were in a 1:1 ratio. The subject’s task was to determine the position (top left, top right, bottom left or bottom right) of the window. The subject pressed the mouse button and communicated the result verbally. Afterwards, the subject was also supposed to determine the window size (small or large), press the button again, and communicate the size determination result verbally. Both head and eye movements were allowed.

We used a balanced within-subjects design. The dependent variables were the correctness of and the required time for localization and size determination and the amount of eye movements and head movements. Eye and head movements were measured in depth image coordinates. For that purpose, we determined the gaze point shift as well as the camera shift between two consecutive depth image video frames. Camera shift or optical flow (which are equal in static environments) are proportional to head motion as the *ANS* system is head-mounted. OpenCV⁵ was used for optical flow calculation from appropriate tracking features. Figure 5 illustrates eye movements and optical flow.

³<http://www.sr-research.com>

⁴<http://www.codemercs.com>

⁵<http://opencv.willowgarage.com/wiki>

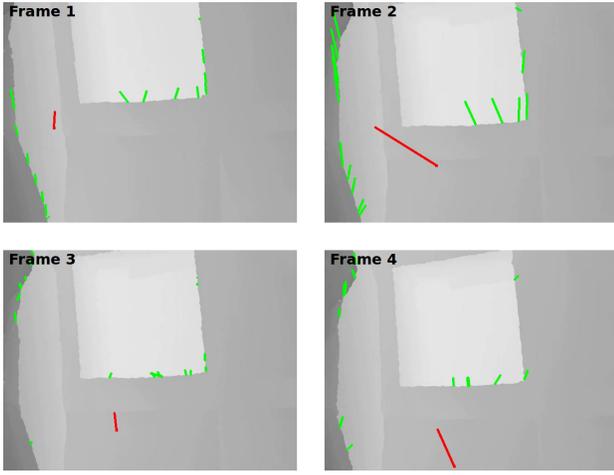


Fig. 5. Eye movements and optical flow in four successive frames of the depth image video. Eye motion velocity vectors are depicted in red. Optical flow vectors of appropriate tracking points are depicted in green. We averaged over the magnitudes of the optical flow vectors in order to estimate the camera motion.

C. Results

The subjects correctly determined the window position in 91.25 percent of the trials which took, on average, 22.31s each. In the window size determination task they succeeded in 71.25 percent of the trials, the average time required being 9.64s. All results are illustrated in Figure 6. For the dependent variables correctness of localization, time required for localization, correctness of size determination and time required for size determination a two-way ANOVA with factors noise and window size has been calculated.

Window size had a significant influence ($F(1,9) = 11.0; p < 0.05$) on the time required for localization, namely 19.59s ($SD = 11.29s$) for large windows and 25.03s ($SD = 14.89s$) for small windows. However, no influence of window size on the localization correctness could be noticed. Size determination correctness was significantly higher ($F(1,9) = 16.4; p < 0.05$) for small windows (83.75 percent, $SD = 21.23$ percent) than for large windows (58.75 percent, $SD = 25.34$ percent). The ANOVA results did not show any significant influence of acoustic noise on the measured values. Also, no interaction effects occurred.

Eye movement velocity during the localization task was 401.3 depth image pixels per second ($SD = 149.69[px/s]$) which corresponds to 36° of visual angle. Significantly less ($t(9) = 2.5; p < 0.05$) eye movements, namely 344.13[px/s] or 31° ($SD = 155.93[px/s]$), were made during the size determination task. The camera motion estimation through optical flow calculation revealed that subjects made considerably less head movements. Head motion velocity also was significantly higher ($t(9) = 2.6; p < 0.05$) during localization (97.87[px/s] or 9° , $SD = 61.49[px/s]$) than during the size detection task (64.04[px/s] or 6° , $SD = 47.68[px/s]$).

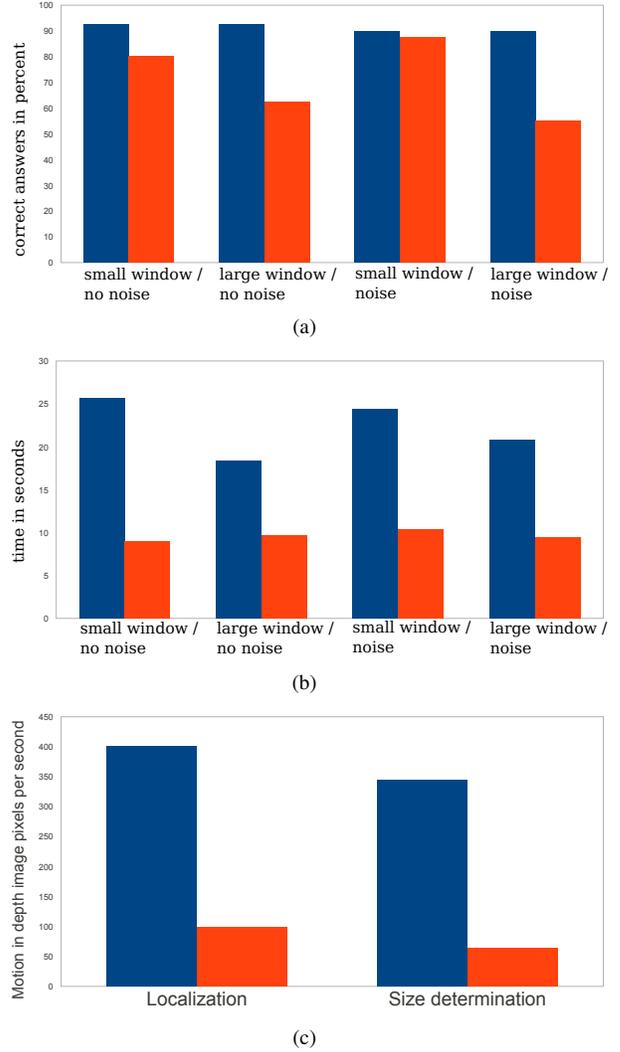


Fig. 6. Results of the proof-of-concept study in the dark. (a) Percentage of correct localizations (blue) and size detections (red) plotted against window size and noise presence. (b) Average time required for localization (blue) and size detection (red) plotted against window size and noise presence. (c) Average eye movement (blue) and head motion (red) velocity plotted against the task. 100 depth image pixels correspond to 9° of visual angle.

D. Discussion

The results of the proof-of-concept study show that the ANS prototype is generally functional. In more than nine out of ten trials localization was successful. Failures can, in many cases, be explained by technical problems such as bad calibration or camera drift. Also, performance differed a lot among the subjects. Size detection seems to be the much more difficult task. Most subjects consciously controlled their gaze for the first time, missing visual feedback complicating it even more. However, experience is expected to greatly improve accuracy.

Large objects were, as expected, localized faster but not better than small objects. At a first glance, it seems surprising that size determination correctness was higher for small windows than for large windows. But "small" was, all in all, called more often than "large", although the experiment design

was balanced. Apparently, objects were perceived smaller than they really were.

Street noise did not have a negative effect on the perception with the *ANS*, at least for our setup where the sonification was clear and simple. A more sophisticated sonification might lead to another result. Also, we did not cross-check. It is possible that the *ANS* sonification disturbs the perception of ambient sounds.

The comparison between head and eye movements shows that the subjects made use of the gaze contingency approach. However, the tasks could also be solved without eye movements. There is no clear evidence that gaze-contingent systems are really faster and more comfortable than static systems. It is plausible that more eye and head movements were needed for localization than for size detection. In the former case, nearly the complete visual field had to be scanned, while in the latter case, the window position was already known.

The darkness experiment has shown that even untrained users can aurally perceive depth information. But it is not known whether gaze-contingent auditory substitution also permits orienting and navigating in unknown environments. We suppose that a lot of practice would be necessary. However, we could not make any studies on this topic because the prototype is not wireless. The user cannot walk around with the helmet put on. For this purpose, a fully mobile version of the system should be developed in the future.

VI. EEG STUDY

A. Objectives and working hypothesis

We were not only interested in the functionality of the system, but also in what happens in the brain when the system is used. Therefore, we studied neural activation in untrained subjects during gaze-contingent audio-visual substitution. The study was mostly exploratory. The results should, therefore, be seen as a starting point for further investigations. Nevertheless, we had the working hypothesis that there is not only auditory perception but also vision-like perceptual and cognitive processing in the *ANS* users' brains.

B. Experimental setup and method

Electroencephalography (EEG) is a method of recording electrical brain activity. Voltage is measured over time by electrodes placed on the subject's scalp. We chose twelve electrode positions from the extended international 10-20 system [28]. The human EEG can be divided into the functional frequency bands delta (< 4 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz) and gamma (> 30 Hz). We only investigated theta, alpha and beta which was split up into beta1 (13-19 Hz) and beta2 (20-30 Hz).

The experiment was conducted with five subjects (four male, one female). There was no real depth measuring in order to have precise control over the stimuli. Instead, stimuli were presented on a screen. First, the measuring electrodes were placed on the subject's scalp and connected to the EEG amplifier. Then, the subject put on the eyetracker which was calibrated and configured to output gaze positions in screen

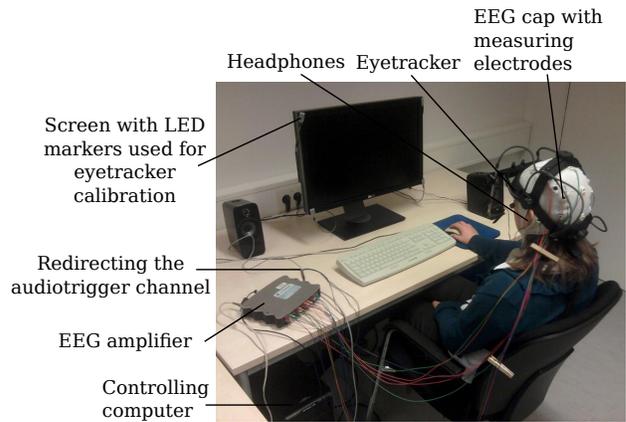


Fig. 7. Experimental setup of the EEG study.

coordinates. This was in contrast to the original *ANS* system which uses head-referenced data. Again, acoustic stimuli were played via earphones (see Figure 7).

There were four experimental conditions, one test and three control conditions, which the subject completed in the following order: visual control condition, auditory control condition, test condition, gaze-contingent control condition. The subject received written and oral instructions and made two try-out runs before each part of the experiment. There were 24 trials per condition. The EEG was recorded during the whole time of the experiment.

- **Test condition:** *All essential elements of object localization and recognition with the ANS: search by eye movements, auditory perception, possibly visual interpretation.* The subject's task was to localize and to detect the shape of an invisible rectangle or triangle (at a size of about 9° of visual angle) on a black screen. If the object was in the direction of gaze, a high pitch (300 Hz) tone was played, otherwise a low pitch (150 Hz) tone was played.
- **Gaze-contingent control condition:** *Isolation of search by eye movements.* The subject's task, again, was to localize and to detect the shape of a rectangle or triangle on the screen. However, there was no sonification. Instead, the subject gaze-controlled a small, round viewing window (at a size of about 3° of visual angle). Thus, when the subject's gaze was close to the target object, the object was partially visible.
- **Auditory control condition:** *Isolation of auditory perception.* The subject listened to a sine tone that alternated between high pitch (300 Hz) and low pitch (150 Hz) for ten seconds per trial. Pitch changes were randomized but were at least one second apart from each other. The subject's task was to attentively listen to the pitch changes.
- **Visual control condition:** *Isolation of visual perception.* The subject watched a white rectangle or triangle displayed on the screen for ten seconds per trial. The object changed its position several times during the trial. The position changes, again, were at least one second apart

from each other. The subject’s task was to attentively observe the position changes.

We were interested in interactions between several brain areas during the different tasks. According to Lubar [29] pacemaker cells in the thalamus influence the local, regional and global resonances in the cortex, and hence the EEG. Thus, the thalamus determines which cortical brain areas work together and at which frequency. Therefore, significant phase synchrony between two brain areas for a certain amount of time is seen as an indication of a common thalamic source. Due to volume conduction [30] the EEG measured at a certain electrode position does not necessarily correspond to the brain area lying underneath. We handled this issue by applying a spatial laplacian filter. This volume conduction estimation, however, is only approximate because of the relatively small number of electrodes.

We calculated a morlet wavelet transformation of the filtered EEG because synchrony analysis is made in the frequency domain. Lachaux et al. [31] have suggested calculating a *Phase Locking Value (PLV)* as a measure for synchrony. The phase difference $\Phi_{xy}(\tau, s)$ between two EEG signals x and y at time τ and at wavelet scale s can be determined from the complex wavelet coefficients. Then, the *PLV* is defined as

$$PLV_{xy}(\tau, F) = \frac{1}{N} \sum_N 1 - \frac{\Phi'_{xy}(\tau, F)}{\pi}, \quad (2)$$

N being the number of trials and $\Phi'_{xy}(\tau, s)$ being the phase difference averaged over all scales of the frequency band F . Thus, we get a *PLV* between 0 and 1 for each point in time τ and each frequency band F , 0 standing for no synchrony between x and y and 1 standing for maximum synchrony. We applied a method to statistically analyze the *PLV* using so called *Surrogate PLVs* which was also proposed by Lachaux et al. [31]. Data segments of one second length starting from stimulus onset (pitch or object position change) were considered for analysis. We assumed interaction between two brain areas if there was statistically significant synchrony for at least 200 *ms*. Significant synchrony for 400 *ms* or more was interpreted as strong interaction.

C. Results

The results of the phase synchrony analysis are plotted topographically in Figure 8. There is one plot for each condition and each frequency band. The phase synchrony durations have been averaged over all subjects.

Interactions can be noticed in all conditions and frequency bands. There are certain patterns regardless of the experimental condition. Generally, there is more interhemispheric synchrony between correspondent electrodes in the low frequency bands than in the high frequency bands. In the beta band, in contrast, there are more links between the frontocentral and lateral electrodes.

It is harder to detect the essential differences between the conditions. What is most noticeable is the fact that the test condition results differ a lot from all control conditions in

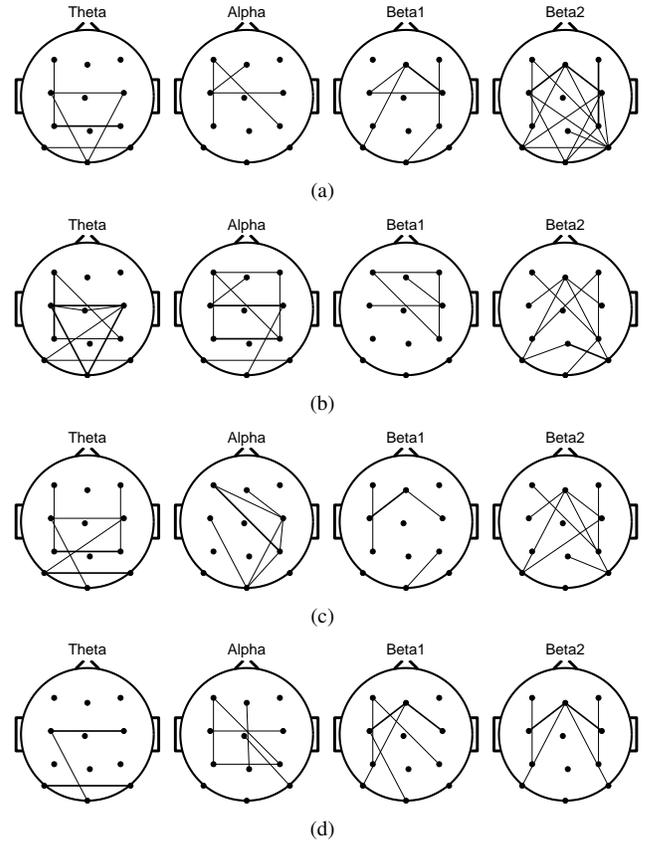


Fig. 8. Topographically plotted phase synchrony. Thin lines: During the regarded period of 1000 *ms* synchrony in the respective frequency band was significant for at least 200 *ms*. Thick lines: Synchrony was significant for at least 400 *ms*. (a) Test condition. (b) Gaze-contingent control condition. (c) Auditory control condition. (d) Visual control condition.

the beta2 band. Firstly, there is generally much more phase synchrony in the test condition. Secondly, there is an especially high number of links between lateral-central and occipital electrodes which is not present in the control conditions.

D. Discussion

Our results give a first impression of cortical interaction during gaze-contingent audio-visual substitution. If our working hypothesis is correct, there should be interaction between auditory and visual brain areas in the test condition. Hence, we should measure phase synchrony between the corresponding electrodes. In this regard, we did not find relevant differences between test and control conditions in the theta, alpha and beta1 frequency bands. The observations in the beta2 band, however, support the hypothesis. From the high number of links between lateral-central (auditory cortex) and occipital (visual cortex) areas we conclude that there may be visual processing of acoustic stimulus changes. Almost no synchrony between neighboring electrodes could be noticed. This might be due to laplacian filtering.

Several studies have shown that synchrony in higher frequency bands, especially in the gamma band, are characteristic for learning processes [32] and multisensory integration [33].

Therefore, it could be worthwhile investigating gamma band synchrony during gaze-contingent sensory substitution. The data from the current experiment, however, are not suited for gamma band analysis because the recordings were made in a standard laboratory without dedicated shielding to prevent net humming artefacts.

VII. CONCLUSION AND FUTURE WORK

The objectives of our current work were to develop and to evaluate an audio-visual substitution system for the blind and visually impaired. The plasticity of the human brain makes it possible for blind people to "see with sound". Eye movement commands serve as spatial indices in mental imagery. The *Auditory Night Sight (ANS)* prototype we have developed realizes the concept of gaze-contingent substitution of spatial vision. We evaluated the system both functionally and neurophysiologically. In a proof-of-concept study in the dark we found that the system is generally functional although there is still room for improvement and enhancement. The findings of the EEG study indicate that gaze-contingent sensory substitution leads to multimodal interaction which is the requirement for neuroplasticity processes.

In the future, the proposed system could help blind and visually impaired people orient and navigate. However, the current version of the system is not wireless. A fully mobile system that allows for navigation experiments ought to be developed. This is possible as depth cameras and eyetrackers are getting smaller and smaller. Then, long term studies should be made in order to investigate learning curves and brain plasticity. Also, a more sophisticated sonification approach might be useful.

Acknowledgment: This work was partially funded by the German Research Council (DFG), grant EXC 277.

REFERENCES

- [1] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, 1992.
- [2] C. Capelle, C. Trullemansa, P. Arno, and C. Veraart, "A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 10, pp. 1279–1293, 1998.
- [3] G. Bologna, B. Deville, and T. Pun, "Sonification of color and depth in a mobility aid for blind people," in *International Conference on Auditory Display, Washington D.C.*, 2010.
- [4] P. Bach-y-Rita, C. C. Collins, F. Saunders, B. White, and L. Scadden, "Vision substitution by tactile image projection," *Nature*, vol. 221, pp. 963–964, 1969.
- [5] R. Farcy and Y. Bellik, "Locomotion assistance for the blind," in *Universal Access and Assistive Technology*. Springer, 2002.
- [6] S. A. Brandt and L. W. Stark, "Spontaneous eye movements during visual imagery reflect the content of the visual scene," *Journal of Cognitive Neuroscience*, vol. 9, no. 1, pp. 27–38, 1997.
- [7] B. Laeng and D. S. Teodorescu, "Eye scanpaths during visual imagery reenact those of perception of the same visual scene," *Cognitive Science*, vol. 26, no. 2, pp. 207–231, 2002.
- [8] R. Johansson, J. Holsanova, and K. Holmqvist, "Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness," *Cognitive Science*, vol. 30, no. 6, pp. 1053–1079, 2006.
- [9] C. V. Parise, C. Spence, and M. O. Ernst, "When correlation implies causation in multisensory integration," *Current Biology*, vol. 22, no. 1, pp. 46–49, 2012.
- [10] D. Alais and D. Burr, "The ventriloquist effect results from near-optimal bimodal integration," *Current Biology*, vol. 14, no. 3, pp. 257–262, 2004.
- [11] N. Doidge, *The Brain That Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science*. Scribe Publications, 2010.
- [12] N. W. Milgram, C. M. MacLeod, and T. L. Petit, *Neuroplasticity, learning, and memory: Proceedings of a symposium held at the University of Toronto, Scarborough, Ontario, March 25, 1986*. Liss, 1987.
- [13] J. Liepert, W. H. R. Miltner, H. Bauder, M. Sommer, C. Dettmers, E. Taub, and C. Weiller, "Motor cortex plasticity during constraint-induced movement therapy in stroke patients," *Neuroscience Letters*, vol. 250, no. 1, pp. 5–8, 1998.
- [14] N. Lessard, M. Paré, F. Lepore, and W. Lassonde, "Early-blind human subjects localize sound sources better than sighted subjects," *Nature*, vol. 395, no. 6699, pp. 278–280, 1998.
- [15] J. A. M. Rojas, J. A. Hermosilla, R. S. Montero, and P. L. L. Espí, "Physical analysis of several organic signals for human echolocation: Oral vacuum pulses," *Acta Acustica united with Acustica*, vol. 95, no. 2, pp. 325–330, 2009.
- [16] L. Thaler, S. R. Arnott, and M. A. Goodale, "Neural correlates of natural human echolocation in early and late blind echolocation experts," *PLoS ONE*, vol. 6, no. 5, p. e20162, 2011.
- [17] N. Vuillerme, N. Pinsault, O. Chenu, J. Demongeot, Y. Payan, and Y. Danilov, "Sensory supplementation system based on electro-tactile tongue biofeedback of head position for balance control," *Neuroscience Letters*, vol. 431, no. 3, pp. 206–210, 2007.
- [18] S. K. Nagel, C. Carl, T. Kringe, R. Martin, and P. König, "Beyond sensory substitution - learning the sixth sense," *Journal of Neural Engineering*, vol. 2, no. 4, pp. R13–R26, 2005.
- [19] P. Troyk, M. Bak, J. Berg, D. Bradley, S. Cogan, R. Erickson, C. Kufta, D. McCreery, E. Schmidt, and V. Towle, "A model for intracortical visual prosthesis research," *Artificial Organs*, vol. 27, no. 11, pp. 1005–1015, 2003.
- [20] P. Bach-y-Rita, *Brain mechanisms in sensory substitution*. Academic Press, 1972, p. 99.
- [21] M. Corbetta, E. Akbudak, T. E. Conturo, A. Z. Snyder, J. M. Ollinger, H. A. Drury, M. R. Linenweber, S. E. Petersen, M. E. Raichle, D. C. V. Essen, and G. L. Shulman, "A common network of functional areas for attention and eye movements," *Neuron*, vol. 21, no. 4, pp. 761–773, 1998.
- [22] M. Zöllner, H.-C. J. S. Huber, and H. Reiterer, "Navi - a proof-of-concept of a mobile navigational aid for visually impaired based on the microsoft kinect," in *Human-Computer Interaction - INTERACT 2011*, vol. 6949, 2011, pp. 584–587.
- [23] A. A. Skavenski and R. M. Steinman, "Control of eye position in the dark," *Vision Research*, vol. 10, no. 2, pp. 193–203, 1970.
- [24] D. Kömpf and H.-F. Piper, "Eye movements and vestibulo-ocular reflex in the blind," *Journal of Neurology*, vol. 234, no. 5, pp. 337–341, 1987.
- [25] J. Battle, E. Mouaddib, and J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: A survey," *Pattern Recognition*, vol. 31, no. 7, pp. 963–982, 1998.
- [26] F. Pece, J. Kautz, and T. Weyrich, "Three depth-camera technologies compared," in *First BEAMING Workshop, Barcelona, 2011*, 2011, p. 9.
- [27] A. T. Duchowski, *Eye tracking methodology: Theory and practice*. Springer, 2007.
- [28] F. Sharbrough, G. E. Chatrian, R. P. Lesser, H. Lüders, M. Nuwer, and T. W. Picton, "American electroencephalographic society guidelines for standard electrode position nomenclature," *Journal of Clinical Neurophysiology*, vol. 8, no. 2, pp. 200–202, 1991.
- [29] J. F. Lubar, "Neocortical dynamics: Implications for understanding the role of neurofeedback and related techniques for the enhancement of attention," *Applied Psychophysiology and Biofeedback*, vol. 22, no. 2, pp. 111–126, 1997.
- [30] P. L. Nunez and R. Srinivasan, *Electric fields of the brain: The neurophysics of EEG*. Oxford University Press, 2006.
- [31] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.
- [32] W. H. Miltner, C. Braun, M. Arnold, H. Witte, and E. Taub, "Coherence of gamma-band eeg activity as a basis for associative learning," *Nature*, vol. 397, no. 6718, pp. 434–436, 1999.
- [33] J. Kaiser and M. J. Naumer, "Cortical oscillations and multisensory interactions in humans," in *Multisensory Object Perception in the Primate Brain*. Springer, 2010.